AN IMAGE-BASED TECHNIQUE FOR 3D BUILDING RECONSTRUCTION USING MULTI-VIEW UAV IMAGES

F. Alidoost*, H. Arefi

School of Surveying and Geospatial Engineering, College of Engineering, University of Tehran, Tehran, Iran –
(falidoost, hossein.arefi)@ut.ac.ir

KEY WORDS: Unmanned Aerial Vehicles (UAVs), Modelling, 3D reconstruction, Mesh, Semi Global Matching (SGM).

ABSTRACT:

Nowadays, with the development of the urban areas, the automatic reconstruction of the buildings, as an important objects of the city complex structures, became a challenging topic in computer vision and photogrammetric researches. In this paper, the capability of multi-view Unmanned Aerial Vehicles (UAVs) images is examined to provide a 3D model of complex building façades using an efficient image-based modelling workflow. The main steps of this work include: pose estimation, point cloud generation, and 3D modelling. After improving the initial values of interior and exterior parameters at first step, an efficient image matching technique such as Semi Global Matching (SGM) is applied on UAV images and a dense point cloud is generated. Then, a mesh model of points is calculated using Delaunay 2.5D triangulation and refined to obtain an accurate model of building. Finally, a texture is assigned to mesh in order to create a realistic 3D model. The resulting model has provided enough details of building based on visual assessment.

1. INTRODUCTION

There are various applications of 3D building modelling such as urban mapping, virtual reality, 3D visualization, and analysing. So far, many methods and techniques are developed in computer vision and photogrammetric domains to reconstruct the structure of buildings which can be categorized based on different aspects of modelling such as level of details (LOD), automation levels (fully or semiautomatic), the input data (3D or 2D), and the 3D modelling approaches (model driven and data driven (Tarsha-Kurdi, et. al., 2007)).

A knowledge based modelling can be established to reconstruct the building façade using both geometric and semantic features. For this, the fusion of terrestrial laser point clouds and close range images helps to improve the completeness, speed and adaptability of modelling (Pu, 2010). A semiautomatic modelling was proposed by (Cheng, et. al., 2011) based on a new dynamic selection strategy and K-means clustering to identify the 3D boundaries of buildings. In this method, the region-growing algorithm based on the RANSAC plane-fitting technique is employed for 3D building model reconstruction with high-correctness, high-completeness, and good geometric accuracy (Cheng, et. al., 2011). Based on the concept of LOD defined by CityGML standard, the LOD1 and LOD2 of buildings can be reconstructed using different Minimum Bounding Rectangle (MBR) and 2D projection based methods, respectively. Then, the results of the reconstruction can be used to enhance the high resolution digital surface models, efficiently (Arefi, et. al., 2012). A fully automatic reconstructions of complex buildings can be achieved using both point cloud and close range images and based on a set of geometric primitives, like planes or cylinders. This method include three steps such as camera orientation, image segmentation, and image-based modelling (Reisner Kollmann, 2013). A model driven approach based on geometric primitives can deliver the 3D models of building roofs by solving an optimization problem. For this, the

combination of LiDAR data and aerial imagery is used to achieve the high accuracy of modelling in both the horizontal and vertical directions (Zhang, et. al., 2014). The formal grammar processing can be used for LOD3 information reconstruction of buildings with regular structures. In this method, a dense point cloud and also procedural rules are derived from terrestrial images to automatic inference of a façade grammar in the architectural style of the observed building façade (Fritsch, et. al., 2014). In this study, a semiautomatic method based on the extracted point cloud from multi view images is investigated to generate the 3D model of building structures' details.

2. PROPOSED METHOD

In this paper, an efficient workflow is presented to reconstruct a 3D model of complex building façades using UAV images (cf. Figure 1). The main steps of this work include:

- Pose estimation,
- Point cloud generation,
- 3D modelling.

2.1 Pose Estimation

In an aerial image acquisition framework, the camera position and orientation (pose) parameters are measured using the on-board Inertial Measurement Unit (IMU) and Global Positioning System (GPS). However, due to existing some shift and drift errors in GPS/IMU units, the measured pose parameters are not sufficiently accurate. According to the proposed workflow (Figure 1), the first step is dedicated to improve the accuracy of pose parameters and to calculate 3D coordinates using some corresponding features of images, ground control points and initial pose parameters in an Automatic Aerial Triangulation (AAT) and Bundle Block Adjustment (BBA) solution.

---

* Corresponding author
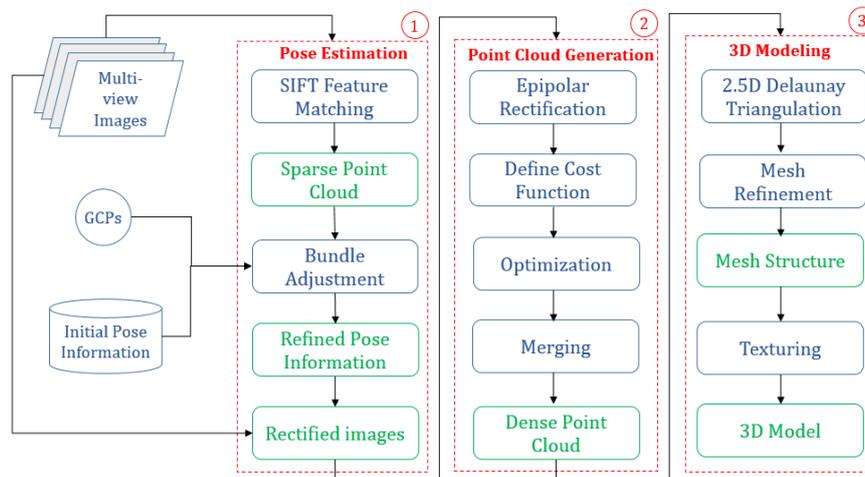
Figure 1. Workflow of 3D Reconstruction

For this, the Scale Invariant Feature Transform (SIFT) (Lowe, 2004) is employed to detect corner points in all images with different scales and orientations which consequently generates sparse point cloud. Next, the 3D coordinates of the given ground control points in WGS84 coordinate system is calculated based on structure from motion algorithms using initial pose parameters as Equation 1.

$$x = K [R | t] X \qquad (1)$$

Where, x is a vector of the known image coordinates and X is a vector of unknown corresponding coordinates in object space. The matrix of K is camera interior parameters and R, t are the pose parameters. The outputs of this step are the refined absolute pose parameters, the oriented images, and the sparse point clouds in a projected coordinate system.

2.2 Point Cloud Generation

After estimating the pose parameters, the range values are calculated from image information of the corresponding pixels in all of images. The effective way to find matched points in different views is to employ epipolar rectification (Zhang, et. al., 1994). In epipolar rectification, all of the images are projected in a 1D space and resampled in order to remove the misalignment in X direction. So, the different between the Y coordinates of corresponding pixels is called parallax in image space which is proportional to the height values in object space. Based on the epipolar matching approach, for a given point in the first image, its corresponding point must lie on the epipolar line in the second view. Therefore, the searching process is along a line rather than the entire image and accordingly causes the matching process to be simpler and faster. Here, the dense stereo matching based on the Semi-Global Matching algorithm (Hirschmueller, 2005) is employed for estimating the disparity between corresponding pixels by minimizing a global cost function. Due to existing high overlap between images, the redundancy of disparity information is used for blunder detection and improving the accuracy of the final depth maps. As the final process, the depth maps from different stereo models are merged together to generate an integrated point cloud.

2.3 3D Modelling

The 3D modelling step is comprising mesh generation and refinement, and texture mapping.

2.3.1 Mesh Refinement: In this paper, the meshing process is utilized in two steps as shown in Figure 1. First, a rough mesh is created based on the Delaunay 2.5D triangulation in order to have a surface covering the whole area to model and avoid holes as much as possible. In order to have a good result of mesh, some parameters should be adjusted such as the size of holes in point cloud to avoid triangulation on gaps and holes, and the average distance between points to have a certain level of details in mesh.

The result of this step is a group of several disconnected surfaces. Therefore, some small meshes should be merged into the biggest parts based on the size of their areas.

To generate a light-weighted and accurate polygonal models, the rough mesh is refined based on the deviation error between the mesh and the estimated best shape. For this, the deviation error parameter is calculated using local curvature (Figure 2).
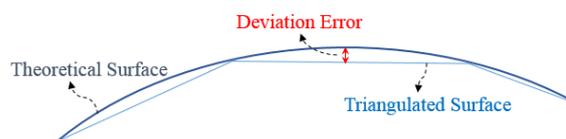

Figure 2. deviation error between the mesh and the estimated best shape

The accuracy of generated mesh depends on the number of triangulations which are defined by deviation error. When the deviation error decreases, the number of points increases. Using the high values of deviation error, the level of detail is low and the resulting surface is roughly defined. Using the small values of deviation error, the resulting surface contains many spikes. So, the best accuracy of mesh is achieved by finding an optimum value of deviation error.

2.3.2 Texture mapping: Finally, in multi-texturing step, the intensity information from multiple pixels of same triangle are combined to produce a unique texture. For this, based on the known internal and external camera parameters of original images from step 1, the texture is assigned to each triangle in 3D model, so the misalignment between textures from different sources could be very small.

3. EXPERIMENTS AND RESULTS

In order to efficiently reconstruct a geometric model, it is often necessary to have some multi-view images from the scene of interest that were taken in a well-designed imaging network. In this paper, two datasets are used. The first dataset is about 80 images acquired using NEX-7_16.0_6000x4000 camera by UAV platform and provided by ISPRS and EuroSDR (Nex, et. al., 2015), is investigated which is captured from a historical building in Dortmund, Germany (cf. Figure 3, left). The second dataset is about 365 images acquired using Olympus PEN-E-P2_17.0_4032x3024 camera by UAV platform over Greifswald Cathedral, St. Nikolai, Germany (cf. Figure 3, right) which is provided by EuroSDR (www.eurosdr.net).

Figure 3. A historical building in Dortmund (left), Greifswald Cathedral, St. Nikolai (right)

Moreover, the metadata of these images including initial pose parameters, the ground control points, and target coordinates are utilized by ISPRS and EuroSDR (Nex, et. al., 2015) which are used as input data in pose estimation step. In feature extraction step, the SIFT descriptors are used to have a fast and efficient matching and as a result, about 15000 feature points are extracted for each image. For first dataset, the mean re-projection error of bundle block adjustment is about 0.1 pixel. Also, the measured spatial resolution of the images is about 0.46 cm. While for the second dataset, these values are about 0.025 m and 0.50 cm, respectively.

As output of step 2, the generated point cloud of first dataset is containing ca. 42 million points (Figure 4, left and up). Here, one side of the building included 6 million points is utilised for modelling. A sampling method based on point distance thresholding is used to have a light version of data (cf. Figure 4, left and down). The threshold parameter is set about 0.01 m to have enough points on some small structures. The second point cloud contains about 7 million points (Figure 4, right) which is small enough for processing.
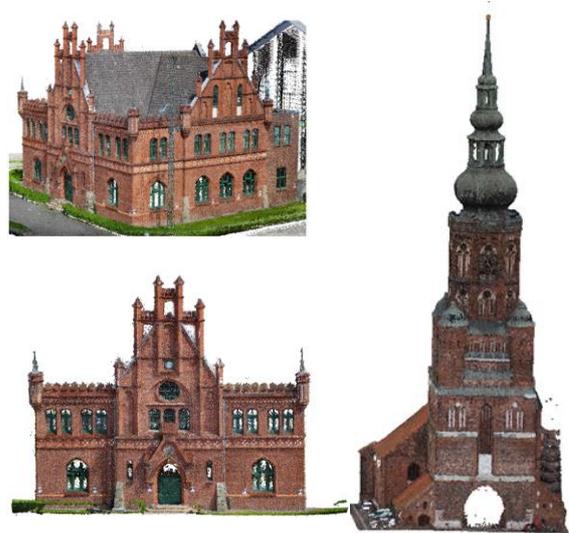
Figure 4. The generated point cloud of first dataset, whole of building (left and up) and the selected side for modelling (left and down), The generated point cloud of second dataset (right)

For mesh generation, the size of the holes in point cloud and the average distance between points are set to 0.05m and 0.1m, respectively for first dataset, and 0.1m and 0.3m for second dataset. The optimum values of these parameters are defined experimentally. Therefore, a Delaunay 2.5D triangulation is applied on point cloud and accordingly a mesh model with about 1 million triangles is generated as shown in Figure 5. In this step, some holes should be filled because of windows or occlusion (for example the holes at top of the door in Figure 4, Left) using curvature continuity constraint.
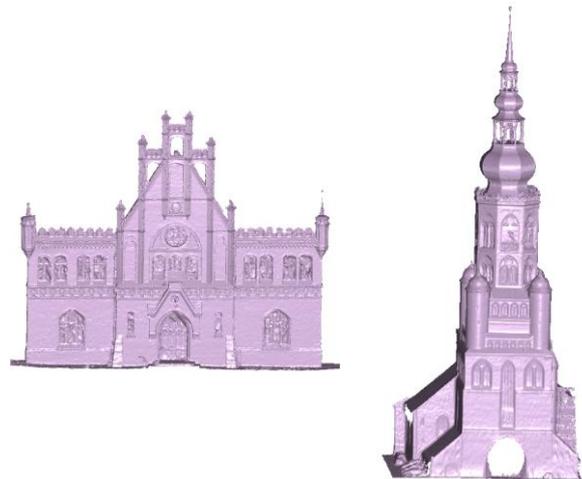

Figure 5. The resulting mesh

As shown in Figure 6, based on estimated values of camera parameters for acquired images, the texture can be assigned to each triangle of mesh. The images for texture mapping should be selected by user and based on illumination conditions to decrease some image distortion. On the other hand, there is no texture information in occluded area and so the quality of resulting texture is decreased in some parts of 3D model.

Figure 6. The results of texture mapping

Based on visual assessment, the generated 3D model and its texture look realistic enough to view the details of the building facades (cf. Figure 5). Moreover, the absolute orientation of the multiple images can improve the texture results and leads to a good seamless model in edges and corners of objects and therefore, the texture transition will be very smooth between different triangles.

4. CONCLUSION

There are many measurement systems or techniques to generate point cloud for 3D modelling such as 3D scanners, digitizers, CMM, laser trackers, laser planes (laser triangulation), time-of-flight lasers (laser-grammetry), photogrammetry or stereo-photogrammetry. The efficient method for 3D modelling is to generate a mesh from point cloud, so the details can be visible from different viewpoints. In this paper, the capability of UAV images is investigated for modelling of urban buildings based on three steps: pose estimation, point cloud generation, and 3D modelling. In this workflow, the interaction of user is limited to initialize some parameters and there is no need to additional dataset such as primary CAD models. Also, the best result of modelling depends on the quality of the original point cloud and user's expectations. Moreover, the large memory requirement due to very dense data is the main problem and critical issue in this work.

REFERENCES

Arefi, H., Reinartz, P., 2012. Multi-level Building Reconstruction for Automatic Enhancement of High Resolution DSMs. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Volume XXXIX-B2, 2012.

Cheng, L., Gong, J., Li, M., Liu, Y., 2011. 3D Building Model Reconstruction from Multi-view Aerial Imagery and Lidar Data. *Photogrammetric Engineering & Remote Sensing*, Vol. 77, No. 2, February 2011, pp. 125–139.

Fritsch, D., Becker, S., Rothermel, M., 2014. Modeling Façade Structures Using Point Clouds From Dense Image Matching. *International Journal of Structural Analysis & Design* – IJSAD, Volume 1, Issue 1.

Hirschmueller, H., 2005. Accurate and efficient stereo processing by semi-global matching and mutual information. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. *IEEE Computer Society Conference on*, Vol. 2, IEEE, pp. 807–814.

Lowe, D., G., 2004. Distinctive Image Features from Scale-Invariant Keypoints. Accepted for publication in the *International Journal of Computer Vision*, 2004.

Nex, F., Gerke, M., Remondino, F., Przybilla H.-J., Bäumker, M., Zurhorst, A., 2015. ISPRS Benchmark for Multi-Platform Photogrammetry. ISPRS Annals of the Photogrammetry, *Remote Sensing and Spatial Information Sciences*, Vol. II-3/W4, pp.135-142.

Pu, S., 2010. Knowledge based building facade reconstruction from laser point clouds and images. Published by: NCG, Nederlandse Commissie voor Geodesie, Netherlands Geodetic Commission, Delft, the Netherlands. ISBN: 978 90 6132 319 3.

Reisner-Kollmann, I., 2013. Reconstruction of 3D Models from Images and Point Clouds with Shape Primitives. Submitted in partial fulfilment of the requirements for the degree of Doktorin, Technische Universität Wien.

Tarsha-Kurdi, F., Landes, T., Grussenmeyer, P., Koehl, M., 2007. Model-driven and data-driven approaches using lidar data: analysis and comparison, *ISPRS Workshop*, *Photogrammetric Image Analysis (PIA07)*, Sep 2007, Munich, Germany. pp.87-92, 2007.

Zhang, Z., Deriche, R., Faugeras, O., Tuan Luong, Q., 1994. A Robust Technique for Matching Two Uncalibrated Images through the Recovery of the Unknown Epipolar Geometry. *Institut National De Recherche En Informatique Et En Automatique*, Programme 4, Robotique, image et vision, Projet Robotvis, Rapport de recherche n 2273, ISSN 0249-6399.

Zhang, W., Wang, H., Chen, Y., Yan, K., Chen, M., 2014. 3D Building Roof Modeling by Optimizing Primitive's Parameters Using Constraints from LiDAR Data and Aerial Imagery. *Remote Sens*. 2014, 6, 8107-8133; doi:10.3390/rs6098107